Asymptotic extremal distribution for non-stationary, strongly-dependent data

Gonzalo Perera 1 and Carolina Crisci 1

1 Polo de Desarrollo Universitario Modelización y Análisis de Recursos Naturales, Centro Universitario Regional del Este, Universidad de la República. Interseccion de Ruta 9 con Ruta 15, Rocha, Uruguay.

1. Abstract

2. Introduction

The statistical analysis of extreme values has a wide and vast domain of applications on many disciplines. Extreme wind speeds are a key input for design in Structural Engineering, maximum levels of traffic are crucial for design and operation in Telecommunications Networks, maximum tides are essential for any policy concerning coast resources management, extreme events on chemical, physical or biological conditions may affect dramatically very sensitive ecosystems, etc.([2],[4],[5],[8],[9],[10],[17]).

The classical Fisher-Tippet-Gnedenko theory determines the asymptotic behavior of the maximum of an iid sequence of random variables, with no other possible non-degenerated limit that the class of extremal distributions ([6],[7]).

Since three decades ago we know that classical theory applies to a stationary and weakly dependent sequence of random variables, thanks to the work of Leadbetter, Lindgren and Rootzen ([12],[13],[14],[18],[19],[20],[21],[22]).

However in many real examples, a very big collection of maximal registers of a large series of measurements, does not fit to any extremal distribution. This often is associated to phenomena where the observed system may assume different states that produces drastic changes and that even may introduce long range dependence on data. Think, for instance, of the classical series of data of the Nile River with its very ancient regime of annual floods, and other large series of hydrological data, in particular those related to the impact of climate change ([11],[15],[23]).

At a theoretical level, we will apply in the extremal context a method that has been used for different purposes in presence of non-stationary and possibly strong dependent data, where a random covariable indicates the state of the system, and the global behavior may be represented by a mixture of models. For instance, that is the case of Compound-Poisson approximation of High-Level Exceedances of time series ([3]), asymptotic behaviour of averages ([16]), and Nadaraya-Watson regression for functional data ([1]).

In this paper we will consider data that depend of two independent components: on one hand, a categorical covariable process that describe the state of the system, that may be neither stationary nor weakly dependent, and that only satisfies that the mean frequency of each state has a (possibly random) limit, and, on the other hand, an iid noise. We will assume that for a given state j on the covariable, the maximal asymptotic distribution is non-degenerated and depends on j. In the main result of the paper we will show that if the covariable process satisfies a weak-dependence assumption, the maximal asymptotic distribution of our data is still an extremal distribution, consistently with Lindgren, Leadbetter and Rootzen results. But we will also show that if this covariable process has a strong dependence structure, then this maximal asssymptotic distribution is no longer an extremal one, but a mixture of extremal distributions.

The paper is organized as follows:

We begin in Section 2 by a very brief summary of classical Fisher-Tippet-Gnedenko theory, in particular the characterization of the Maximal Domains of Attraction of extremal distributions and an elementary Lemma that we will use later.

We then present in Section 3 our model, its hypotheses, some examples and the main results with some corollaries.

In section 4 we will fit some simulated strong dependent data to a mixture of extremal distributions, showing that they do not fit extremal distributions. In particular, we will show the impact on return times and return values of missfitting data to a single extremal distribution.

Finally in section 5, we present the conclusions and some further work in progress.

3. Brief review of classical extreme values theory

The classical Fisher-Tippet-Gnedenko theory states that a distribution F belongs to the Maximal Domain of Attraction of an extremal distribution (Weibull, Gumbel, Frechet) H if there exists an iid sequence X_1,\ldots,X_n of random variables with distribution F such that for some deterministic sequences d_n and c_n > 0, we have that
Recall that \( M_F = \sup \{ t \in \mathbb{R} : F(t) < 1 \} \), \( P^{-1}(p) = \inf \{ x \in \mathbb{R} : F(x) \geq p \} \) \( \forall p \in (0, 1) \).

Then the three Maximal Domains of Attraction (MDA) are fully described.

1. Frechet of order \( \alpha : F \) belongs to the MAD\( (\Phi) \), where \( \Phi(x) = \exp\{-x^{-\alpha}\} \forall x > 0 \) and \( \alpha > 0 \) is called the order parameter, if and only if, \( M_F = +\infty \) and for \( x \) tending to infinity, \( 1 - F(x) = \frac{L(x)}{x^\alpha} \) for some slowly varying function \( L \). In that case, in Req.1, the deterministic sequences are \( d_n = 0, c_n = n^{1/\alpha} \).

2. Weibull of order \( \alpha : F \) belongs to the MAD\( (\Psi) \), where \( \Psi(x) = \exp\{-(-x)^\alpha\} \forall x < 0 \), if and only if \( M_F < \infty \) and for \( x \) tending to \( M_F^- \), \( 1 - F(x) = (M_F - x)^\alpha L\left(\frac{1}{M_F - x}\right) \) with \( L \) a slowly varying function. In this case \( \alpha_n = M_F \) and \( c_n = n^{-\alpha} \).

3. Gumbel: \( F \) belongs to the MAD\( (\Lambda) \), where \( \Lambda(x) = \exp\{-e^{-x}\} \forall x \in \mathbb{R} \) iid and only if there exist an \( a < M_F \) (which may be finite or infinite), some \( c > 0 \) and a positive function \( h \), with density \( h' \), such that \( \lim_{x \rightarrow M_F^-} h'(x) = 0 \) and, for \( x \) tending to \( M_F^- \), \( 1 - F(x) \) is equivalent to \( c \exp\{-\int_x^\infty \frac{1}{h(t)} dt\} \), and \( d_n = P^{-1}(1 - 1/n), c_n = h(d_n) \).

**Lemma 1:**

If we denote by \( c_n(\Phi), c_n(\Psi), d_n(\Psi), c_n(\Lambda), d_n(\Lambda) \), the deterministic sequences corresponding to each MDA, we then have

\[
\begin{align*}
&\text{i. if } \alpha_1 < \alpha_2, \quad \frac{c_n(\Phi)}{c_n(\Psi)} \rightarrow n \ 0, \quad \frac{d_n(\Psi)}{c_n(\Psi)} \rightarrow n \ 0 \forall \alpha > 0, \quad \frac{\alpha_n(\Psi)}{c_n(\Psi)} \rightarrow n \ 0 \forall \alpha > 0, \quad \frac{c_n(\Lambda)}{c_n(\Psi)} \rightarrow n \ 0, \quad \frac{d_n(\Lambda)}{c_n(\Psi)} \rightarrow n \ 0 \\
&\text{ii. if } \alpha_1 < \alpha_2, \quad \frac{c_n(\Psi)}{c_n(\Phi)} \rightarrow n \ 0, \quad \frac{d_n(\Psi)}{c_n(\Phi)} \rightarrow n \ 0 \\
&\text{iii. if } \alpha_1 < \alpha_2, \quad \frac{c_n(\Phi)}{c_n(\Psi)} \rightarrow n \ 0, \quad \frac{d_n(\Phi)}{c_n(\Psi)} \rightarrow n \ 0.
\end{align*}
\]

4. **Main Results**

We will assume that the process \( Y \) satisfies (H1). For any state \( j = 1, \ldots, k \), there exists a (possibly random) \( b_j > 0 \) such that

\[
\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^{n} 1\{Y_{i-j} = b_j \} = b_j \text{ a.s.}
\]

If \( I(t) = \sigma \{ Y_i : i \geq t \} \), and \( I(\infty) = \bigcap_{n>1} I(t) \) , then, since \( \forall j \ b_j \) is \( I(\infty) \)-measurable, are deterministic, but, if \( I(\infty) \) is not trivial, \( b_j \) may be non-deterministic, corresponding to strong dependence on the process \( Y \). Let us show this in a very simple case.

**Example 1**

Let \( U \) be a random variable such that \( P(U = 1) = p, P(U = 2) = 1 - p \). Let \( \sigma_1, \ldots, \sigma_n \), an iid sequence of random variables on \( \{1, 2\} \) independent of \( U \) such that \( P(\sigma_1(1) = 1) = \delta, P(\sigma_1(1) = 2) = 1 - \delta \), \( P(\sigma_2(1) = 1) = \eta, P(\sigma_2(1) = 2) = 1 - \eta \), \( \delta, \eta > 0 \), \( \delta < 1, 0 < \eta < 1 \).

Thus, \( \frac{1}{n} \sum_{i=1}^{n} 1\{Y_i = 1\}/U = 1 \) has the same distribution than \( \frac{1}{n} \sum_{i=1}^{n} 1\{\sigma_1(1) = 1\} \xrightarrow{a.s.} P(\sigma_1(1) = 1) \) by the Strong Law of Large Numbers \( = \delta \)

On the other hand \( \frac{1}{n} \sum_{i=1}^{n} 1\{Y_i(1) = 1\}/U = 2 \) has the same distribution than \( \frac{1}{n} \sum_{i=1}^{n} 1\{\sigma_2(1) = 1\} \xrightarrow{a.s.} P(\sigma_2(1) = 1) = \eta \)

Therefore if we assume that \( \delta \neq \eta \), we have that

\[
b_1 = \begin{cases} 
\delta & \text{if } U = 1 \\
\eta & \text{if } U = 2
\end{cases}
\]

\( b_1 \) is not-deterministic and \( I(\infty) \) is not trivial. Similar treatment applies to \( b_2 \). Indeed, for the sake of simplicity we will assume (H2) \( \forall b_j \) may only assume a finite number of values.

Our data will be of the form:

\[
X_i = f(\varepsilon_i, Y_i)(\#eq : 2)
\]

where \( f \) is unknown we will assume (H3)

i. \( \varepsilon_1, \ldots, \varepsilon_n, \ldots \) iid
ii. $Y_1, \ldots, Y_n$ satisfies (H1)

iii. The processes $\varepsilon_1, \ldots, \varepsilon_n$, and $Y_1, \ldots, Y_n$, are independent among them.

Finally we will assume that the state of the system is observable (speach aca o antes al introducir $Y$) and that it affects the extremal behavior of our data.

More precisely, we shall assume:

H4) There exists an integer $f, 1 < f < k$ and an integer $g, 1 < g, f + g < k$ such that

a. For $j = 1, \ldots, f$ the iid process $f(\varepsilon_1, j), \ldots, f(\varepsilon_n, j)$ belongs to MDA($\Phi^i_{\alpha}$),

b. For $j = f + 1, \ldots, f + g$ the iid process $f(\varepsilon_1, j), \ldots, f(\varepsilon_n, j)$ belongs to MDA(A)

c. For $j = f + g + 1, \ldots, k$, the iid process $f(\varepsilon_1, j), \ldots, f(\varepsilon_n, j)$ belongs to MDA($\Psi_{\alpha_2}$),

with $0 < \alpha_{f+g+1} < \ldots < \alpha_k$.

Then we have the following result.

**Theorem 1:**

Under (H1), (H2), (H3) there exists a random variable $Z / \frac{\max(X_1, \ldots, X_n)}{n^{1/\alpha}} \overset{w}{\rightarrow} Z$.

In addition

a. If $I(\infty)$ is trivial, then, the distribution of $Z$ is

$$F_z(x) = \sum_{i=1}^{k} p_i = \Phi_{\alpha_1} \frac{x}{\beta_i}$$

b. If $I(\infty)$ is not trivial and $b_1$ assumes the distribution of $Z$ is

$$F_z(x) = \sum_{i=1}^{k} \Phi_{\alpha_n} \frac{x}{\beta_i}$$

**Proof:**

$$P(\frac{\max(X_1, \ldots, X_n)}{n^{1/\alpha}} \leq x) = \int_{1 \leq k} P(\frac{\max(X_1, \ldots, X_n)}{n^{1/\alpha}} \leq x) / Y_1 = j_1, \ldots, Y_n = j_n dP_{\psi}(j_1, \ldots, j_n, \ldots) (#eq:3)$$

We will re-order the maximum according to the state $j$ taken by each $Y_i$.

Then #eq:3 equals to

$$\int_{1 \leq k} P(\frac{\max\{\max(X_{1_i}, \ldots, X_{1_i})\}_{n^{1/\alpha}}}{n^{1/\alpha}} 1\{Y_1 = \ldots = Y_{n_{1_i}} = 1\}, \ldots, \frac{\max\{X_{n_i}, \ldots, X_{n_i}\}_{n^{1/\alpha}}}{n^{1/\alpha}} 1\{Y_1, \ldots, Y_{n_i} = k\}) \leq \frac{1}{Y} Y = j_1, \ldots, Y_n = j_n dP_{\psi}(j_1, \ldots, j_n, \ldots) (#eq:4)$$

Among the blocks of maximum values taken with $Y$ fixed on a given state, the MDA is different, and taking into account Lemma 1, and (H4), all the blocks after the first, tends in probability to zero, and therefore, the limit as $n$ tends to infinity of #eq:4 is the same than

$$\lim_{n} \int_{1 \leq k} P(\frac{\max\{f(\varepsilon_{1_i}, 1), \ldots, f(\varepsilon_{n_i}, 1)\}}{n^{1/\alpha}}) \leq x / Y_1 = j_1, \ldots, Y_n = j_n dP_{\psi}(j_1, \ldots, j_n, \ldots) (#eq:5)$$

Using that $\varepsilon_i$ are iid, the previous limit is the same than

$$\lim_{n} \int_{1 \leq k} P(\frac{\max\{f(\varepsilon_{1}, 1), f(\varepsilon_{1}, 1)\}_{n^{1/\alpha}}} / \sum_{i=1}^{n} 1\{Y_i = 1\} = n_1) dP_{\psi}(j_1, \ldots, j_n) (#eq:6)$$

Since $\frac{1}{n} \sum_{i=1}^{n} 1\{Y_i = 1\} \overset{a.s.}{\rightarrow} b_1$ and by Dominated Convergence Theorem the limit in (#eq:6) equals to

$$\int_{0}^{1} \lim_{n} \frac{P(\frac{\max\{f(\varepsilon_{1}, 1), \ldots, f(b_1 n, 1)\}}{(b_1 n)^{1/\alpha}}) \leq x / b_1 dP_{\psi}(u) (#eq:7)$$

Therefore if $I(\infty)$ is trivial, $b_1$ is deterministic and (#eq:7) equals to $P(b_1^{1/\alpha} \Gamma \leq X)$, with $\Gamma \sim \Phi_{\alpha_1}$ and part a) of Theorem 1 follows.
If $b_1$ is random using (H2) then part b) of Theorem 1 follows.

**Remark:**
It is clear that (H2) may be removed, leading in (b) to an integral with respect to the distribution of $b_1$ instead of a sum.

**Example 2**
Following the ideas of Example 1, consider $\sigma(1), \sigma(2)$ independent such that $P(\sigma(1) = 1) = \delta, P(\sigma(1) = 2) = 1 - \delta$.

$P(\sigma(2) = 1) = \eta, P(\sigma(2) = 2) = 1 - \eta, 0 < \delta < 1, 0 < \eta < 1$ and $\delta \neq \eta$

Taking $\sigma_1(1), \sigma_1(2), \ldots, \sigma_n(1), \sigma_n(2), \ldots$ a sequence of independent copies of $(\sigma(1), \sigma(2))$ it turns out that if $U$ is a fixed random variable such that $P(U = 1) = p, P(U = 2) = 1 - p, 0 < p < 1$ then if $Y_i = \sigma_i(U)$, we have that $Y_1, \ldots Y_n, \ldots$ fulfills (H1),(H2) with $b_1, b_2$ random variables such that

$$b_1 = \begin{cases} \delta & \text{if } U = 1 \\ \eta & \text{if } U = 2 \end{cases}$$

and

$$b_2 = \begin{cases} 1 - \delta & \text{if } U = 1 \\ 1 - \eta & \text{if } U = 2 \end{cases}$$

Thus if we assume $0 < \alpha_1 < \alpha_2$ and consider two independent sequences $V_1^{(1)}, \ldots, V_n^{(1)}, \ldots, iid \sim F^{(1)}, V_1^{(2)}, \ldots, V_n^{(2)}, \ldots, iid \sim F^{(2)}$ with

$F^{(i)} \in MDA(\Phi_{\alpha_i})$ and we set:

- a. If $\sigma_i(U) = 1, X_i = V_i^{(1)}$
- b. If $\sigma_i(U) = 2, X_i = V_i^{(2)}$

then, by Corollary 1,

$$\max(X_1, \ldots, X_n) \xrightarrow{w} n \Phi_{\alpha_1} \left( \frac{1}{\sqrt{n}} \right) + (1 - p) \Phi_{\alpha_2} \left( \frac{1}{\sqrt{n}} \right),$$

a mixture of Frechet distributions.

Figure 1 shows the difference between a Frechet model with $\alpha_1 = 1$ (F1), and a mixture with the same $\alpha_1$ (F1:Mixture), $p = 0.25$; $\delta = 0.20 ; \eta = 0.80$.

Even for such moderated values, mixtures models is less pesimistic , and for instance considering the value $u = 1.90$ for mixture model the probability of exceeding $u$ is 0.283 and its return time 2.444 (30.9% more frequent). Off course with other settings of parameters differences may be stronger and on the contrary sense (underestimation of simpler models). The aim here is to show than even in not so-complex situations, ignoring the real complexity may lead to significative errors.
Example 3

We will now simulate data leading to the mixture model of the preceding example and see if even in such a case where differences with a simple Frechet model it is not dramatical, we are able to show which is the most suitable model. If we consider a distribution function of the form:

\[
F(x) = 1 - \frac{1}{x} \forall x > 0, (\#eq: 8)
\]

then \(P_1(x) = \frac{1}{x}L(x),\) with \(L(x) = \frac{x_{\alpha}}{1 + x}\) which is a slowly varying function and therefore a \(F^{(c)} \in MDA(\Phi_\alpha).\)

Its inverse function is

\[
P^{(c)}(a) = \frac{1}{1 - a} - 1 \forall y \in (0, 1)(\#eq: 9)
\]

Consider \(n\) large (say \(n = 500\)) and \(N\) very large (say \(N = 2000\)) and consider \((\sigma_1, \sigma_2)\) a random variable such that \(\sigma_1, \sigma_2\) independent

\[
P(\sigma_1 = 1) = \delta, P(\sigma_1 = 2) = 1 - \delta, P(\sigma_2 = 1) = \eta, P(\sigma_2 = 2) = 1 - \eta, \text{ with } 0 < \delta < 1, 0 < \eta < 1, \delta \neq \eta.
\]

For instance, we shall take \(\delta = 0.20; \eta = 0.80.
\]

Consider \((\sigma_{1,1}, \sigma_{1,2}, \ldots, \sigma_{2,1}, \sigma_{2,2}, \ldots, \sigma_{2,n})\) a matrix of independent copies of \((\sigma_1, \sigma_2).\) As seen in Example 1, if for each \(i, j\) we define \(Y_{i,j} = \sigma_{i,j}U_j\)

Consider \(U_1, \ldots, U_N\ iid\) such that \(P(U_j = 1) = p, P(U_j = 2) = 1 - p\) and take \(p = 0.25.\)

Then, for each fixed \(i Y_{i,1}, \ldots, Y_{i,n}\) fulfills (H1),(H2)

If we consider two independents \(i, j\) matrix \(e_{ij}^{(1)}(\text{with } 1 \leq i \leq N\) and \(1 \leq j \leq n)\) and \(e_{ij}^{(2)}(\text{with } 1 \leq i \leq N\) and \(1 \leq j \leq n)\) such that \(e_{ij}^{(1)} \sim F^{(1)}(\text{for } \alpha = 1), e_{ij}^{(2)} \sim F^{(2)}(\text{for } \alpha = 2)\)

and we finally set for each \(i = 1, \ldots, N,\)

\[
X_{i,j} = \begin{cases} e_{ij}^{(1)} & \text{if } \sigma_{i,j}(U_i) = 1 \\ e_{ij}^{(2)} & \text{if } \sigma_{i,j}(U_i) = 2 \\ \end{cases}
\]

then by Corollary 1, \(M_i = \frac{\max(X_{i,1}, \ldots, X_{i,n})}{N}\) must be close to distribution \(MF(x) = 0.25\Phi_{\alpha 1/\xi} + 0.75\Phi_{\alpha 1/\xi}^8\)

We will now compute the historam of \(M_1, \ldots, M_N\) and see if it fits better to MF or to a classical \(\Phi_1\) model.

Conclusions

References


